ORBITA: Jurnal Pendidikan dan Ilmu Fisika
ISSN : 2460-9587 (print) | ISSN : 2614-7017 (elektronik)
Volume 11, No. 2, November 2025, pp. 82-91

Integration of magnus thermodynamic parameters and machine learning algorithms in rainfall prediction

Ayu Aprilia[1*], Hanifah Zakiya[2], Gurum Ahmad Pauzi[1], Amir Supriyanto[1], Syafriadi[1]

[1]Department of Physics, University of Lampung, Lampung, Indonesia
[2]Department of Physics Education, University of Lampung, Lampung, Indonesia
*Correspondence: ayu.aprilia@fmipa.unila.ac.id

Received: 02 September 2025 | Revised: 21 November 2025 | Accepted: 23 November 2025 | Published Online: 30 November 2025

Abstract

Atmospheric physics is very useful in predicting rainfall, particularly for analyzing air saturation conditions as a prerequisite for condensation. This study aims to model rainfall prediction using thermodynamic parameters, namely relative humidity (RH) and dew point temperature difference (ΔT). These parameters were collected from BMKG Lampung meteorological data (2022–2024) and processed using the Magnus equation. ΔT is important as a sensitive indicator of air unsaturation. The data were statistically analyzed and modeled using a Gradient Boosting Classifier. The results obtained indicate a strong correlation between RH and ΔT and rainfall events (point-biserial correlation of 0.475). Furthermore, ΔT during rainfall is lower (average 2.87°C) and stable, indicating near-saturation conditions. During the evaluation stage, the model achieved 76% accuracy and 84% recall during rainfall. The model's good performance proves the effectiveness of physical parameters as predictive features. Finally, the model was implemented in a Flask-based web application for practical accessibility.

Keywords: atmospheric; dew point; magnus equations; rain prediction; gradient boosting.

How to Cite: Aprilia, A., Zakiya, H., Pauzi G. A., Supriyanto, A., Syafriadi. (2025). Integration of magnus thermodynamic parameters and machine learning algorithms in rainfall prediction. *ORBITA: Jurnal Pendidikan dan Ilmu Fisika, 11*(2), 82-91. https://doi.org/10.31764/orbita.v11i2.34505

INTRODUCTION

The formation of rain is fascinating subject of study for atmospheric physics. Rain is a complex phenomenon that involves thermodynamic processes, radiation, and fluid dynamics (Andrews, 2010). Precipitation is a prerequisite for rain that is greatly influenced by water saturation. Meteorological parameters such as temperature, relative humidity, dew point, and wind speed are variables that determine water saturation (Catursari, 2024). In atmospheric review, the formation of rain involves both stochastic, deterministic processes, as well as the microphysics of precipitation, which mutually influence the development of a dynamic weather system (W. Wang & Hocke, 2022).

Research conducted in the Tibetan Plateau has yielded positive results regarding the relationship between extreme rainfall and surface temperature (about 2.5%/°C) and dew point (about 3.5%/°C). These relationships indicate that rainfall is influenced by thermodynamic conditions such as temperature

(H. Wang et al., 2022). In the context of global climate change, rainfall patterns have undergone significant changes. Rainfall patterns have become increasingly variable to the point of being difficult to predict (Arnah Ritonga et al., 2025). If climate change due to global warming continues, it is estimated that two-thirds of the Earth's land area will experience increased rainfall. Increased rainfall occurs due to thermodynamic effects (warm air holds more water vapor) and dynamic effects (atmospheric circulation patterns) (Schneider et al., 2010).

The phenomenon of global warming presents a challenge for mitigating hydrometeorological disaster, such as flooding and drought. Therefore, there is a need for more adaptive and accurate rain prediction models and methods. Traditional Numerical Weather Prediction (NWP) methods have limitations in accuracy, especially for local-scale precipitation (Huertas-Abril & Palacios-Hidalgo, 2023). To address this problem, many researchers are developing machine learning approaches. Machine learning approaches have proven effective in predicting various phenomena in human life, from earthquake predictions to coffee yield forecasts (Aprilia, Wahidin, & Abdurrahman, 2025; Aprilia, Wahidin, & Syafriadi, 2025).

Research on the use of machine learning to predict rain has been conducted using the random forest method. The random forest model is capable of achieving 90% accuracy with an AUC of 0.904 (Chen et al., 2025). Another study on hybrid approaches that combine machine learning with traditional NWP methods increased accuracy by 25%. The use of machines has challenges in determining relevant input variables (Markovics & Mayer, 2022). If predictions only use the most important variables, performance is better than using all variables.

In Indonesia, in 2006, BMKG transitioned from manual measurements to Automatic Weather Stations (AWS) (Halida et al., 2024). This transition has provided a positive correlation, thereby strengthening the consistency and quality of weather data. Additionally, BMKG has integrated a monitoring system with 21 radars across Indonesia. BMKG transitioned to support the management of real-time data (Prakasa & Utami, 2019). The quality and consistency of the data serve as the foundation for implementing rain prediction based on machine learning using meteorological variables.

Atmospheric physics is a fundamental science in predicting rainfall. This research will develop a rainfall prediction model by processing meteorological variables. This study focuses on the Lampung region, where Lampung is a province with agricultural and plantation commodities. When rainfall patterns cannot be predicted, it will threaten agricultural and plantation yields. This research utilizes data from four BMKG observation stations in Lampung. The prediction model is based on a Gradient Boosting Classifier using historical data from BMKG. This study offers a new approach to optimize rainfall prediction. Rainfall prediction is correlated with thermodynamic variables such as temperature gradient and relative humidity. Furthermore, to make the model more applicable, it is developed into a web platform. This is expected to strengthen local climate resilience across various sectors.

METHODS

This research employs a quantitative approach, utilizing descriptive statistical analysis and correlation tests. Furthermore, the data is analyzed using a machine learning approach, and the results of these tests are presented on a website to enhance the research's applicability. The data used comes from the Meteorology, Climatology, and Geophysics Agency (BMKG). BMKG data can be accessed at https://dataonline.bmkg.go.id/data-harian. This study focuses on Lampung Province, so the data are collected from four UPT BMKG Lampung Stations. The distribution of the UPT BMKG Lampung stations is presented in Table 1.

Table 1. Location of UPT BMKG station

Location	Station Name	Latitude	Longitude
UPT South Lampung	Radin Inten II Meteorological Station	-5.16000	105.11000
UPT North Lampung	North Lampung Geophysical Station	-4.83631	104.87000
UPT Pesawaran	Lampung Climatology Station	-5.17236	105.18000
UPT Bandar Lampung	Panjang Maritime Meteorology Station	-5.17236	105.18000

Table 2. Variables and Unit

Variables	Unit
Rainfall (RAIN)	Ml
Temperature (TEMP)	^{o}C
Relative Humidity (HUM)	%
Wind Speed (WIND)	m/s

The data used in this study is daily data from January 1, 2022, to December 31, 2024. The initial variables used are taken from previous research as presented in Table 2. Data processing and the testing of machine learning models are done using Python (Lakshmi, 2018). This research uses Python 3.10 with the main libraries cikit-learn 1.4, NumPy 1.26, Pandas 2.2, and Matplotlib 3.8, which are run through Jupyter Notebook in Anaconda Distribution 2024. Data processing is carried out according to the flowchart in Figure 1.

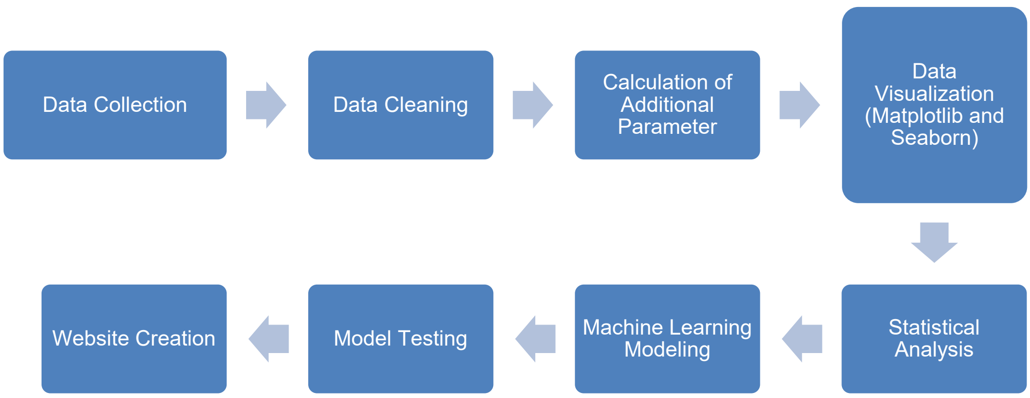

Figure 1. Research Flow

The initial stage involves data cleaning aimed at removing missing values. In the data cleaning step, the data will be collected in its entirety and converted into numerical form (pd.to_numeric), and then empty values (NaN) will be removed (Ningsih et al., 2024). Next, the data is grouped based on whether it is raining or not by creating a binary rain status column with a value of 1 if RAIN ≥ 0.1 mm and zero if it is not raining (Arnah Ritonga et al., 2025). The next stage involves calculating additional parameters, specifically the saturated vapor pressure (e_s) and actual vapor pressure (e), using the Magnus equation.

The equation used to calculate the saturated vapor pressure (e_s) is given by equation 1 and equation2 (W. Wang et al., 2011).

$$e_s = 6.112 \ \exp\frac{17.62 \times T}{243.12 + T} \tag{1}$$

$$e = \frac{RH}{100} \times \ e_s \tag{2}$$

The dew point (T_d) is calculated using the Magnus equation with values a=17.62 and b=243.12 as follows (Magnus) is given by equation 3 and equation 4 (Alduchov & Eskridge, 1988). The results obtained are then visualized using Matplotlib and Seaborn.

$$T_d = \frac{b \times \alpha}{a - \alpha} \tag{3}$$

$$\alpha = \ln\frac{RH}{100} + \frac{a \times T}{b + T} \tag{4}$$

The statistical analysis test in this research used Point-Biserial Correlation to examine the relationship between RHUM and rainfall occurrences as a binary variable. A t-test was conducted to see how the average wind speed differs under saturated and unsaturated air conditions. The results are also complemented by descriptive statistical analysis, including mean, median, minimum value, maximum value, and standard deviation. The processing and analysis utilized libraries including Pandas, NumPy, Matplotlib, Seaborn, SciPy, scikit-learn, and joblib.

The machine learning model used in this research employs the Gradient Boosting Classifier algorithm. The modeling uses an 80% training data and 20% testing data split. Model development was selected using the Gradient Boosting Classifier method with parameter optimization (learning rate, number of estimators, and maximum depth) with GridSearchCV to obtain the best performance. The trained model will be calibrated with probabilities using CalibratedClassifierCV. This step will improve the consistency of predictions, especially for light and moderate rainfall classification. The final model will be exported to .pkl format with the joblib library. After the modeling is completed, the model is evaluated using a Classification report (accuracy, precision, recall, F1-score) and a confusion matrix (Hikmayanti Handayani et al., 2025). Model evaluation is conducted to determine the percentage accuracy of the created model.

The final stage involves creating a simple web application using the Flask framework. The model will be made into a web application using Flask 3.0. In the backend, the application receives user input in the form of temperature, humidity, and wind speed. Then, the input must be validated with the final model that has been prepared beforehand. Then, the prediction results and scores will be sent to the web interface using HTML, CSS, and JavaScript. Inference is made in real time with an average response time of less than 0.2 seconds. This is expected to be accessible to general users without the need for special computing devices.

RESULTS AND DISCUSSION

Relative humidity (RH) is a key parameter to indicate the amount of water vapor present in the air compared to the maximum amount the air can hold at a specific temperature. Its value is influenced by the relationship between air temperature and saturation vapor pressure (e_s), as shown in Equation (1). Through this relationship, temperature changes directly affect the actual vapor pressure (e) and, consequently, the relative humidity, which plays a crucial role in atmospheric processes such as cloud formation and precipitation.

The results are displayed in a box plot to see the variation in humidity during rain and when it is not raining. Based on the boxplot, it can be seen how the distribution of relative humidity compares under rainy and non-rainy conditions. Figure 2 shows that relative humidity is higher during rain. The variation in relative humidity during non-rainy periods is greater, as evidenced by the longer boxplot. Conversely, humidity during rain tends to be stable and high.

The results in image 2 are consistent with the theory that water vapor (H_2O) is the main material for the formation of clouds and rain. Relative Humidity (RH) is a measure of how saturated the air is with water vapor compared to its maximum capacity (Andrews, 2010). The raindrops formed are the result of the condensation of water vapor on condensation nuclei. The condensation process is very effective when the air reaches the saturation point of RH=100%. This theory is supported by a correlation value of 0.475 between RH and the occurrence of rain. The positive results indicate that the higher the relative humidity, the greater the likelihood of rain.

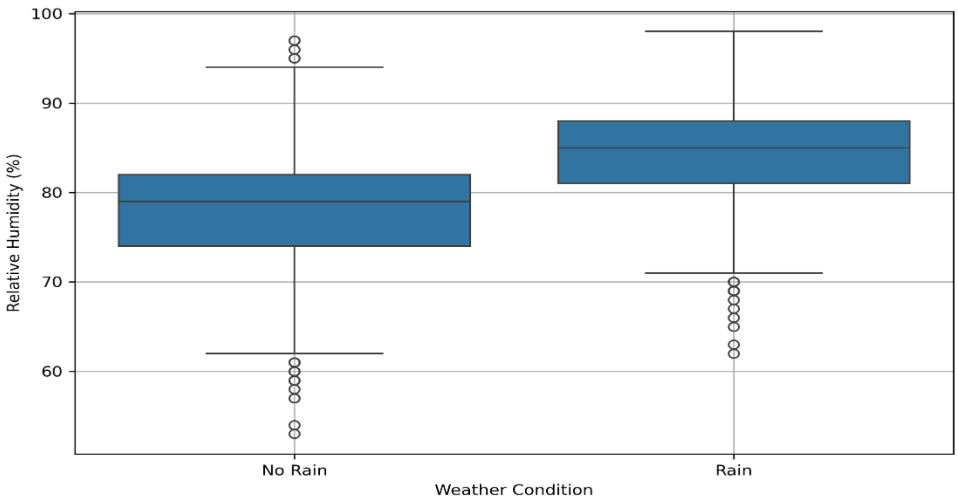

Figure 2. Relative Humidity Distribution during Rain vs No Rain

Figure 3 illustrates the distribution of relative humidity in relation to rainfall and its corresponding frequencies. The results indicate that the scatter plot points form a positive pattern. The results are evident when rainfall tends to increase, then RH rises, although there appears to be variation at RH > 90. The histogram reinforces this by showing that some rainfall occurs with humidity concentrations in the range of 80% to 95%. The peak of the histogram is at a frequency of 85% to 90%. The results obtained are in accordance with the theory of rain formation through the processes of adiabatic cooling and condensation, as physically observed. The air must be in a saturated condition for the vapors to condense into raindrops. The condensation process occurs in the atmosphere, as indicated by high relative humidity values (Andrews, 2010). An anomalous state can occur when humidity is present, but there is no rain due to the absence of convection energy, air stability, and cloud droplet coalescence.

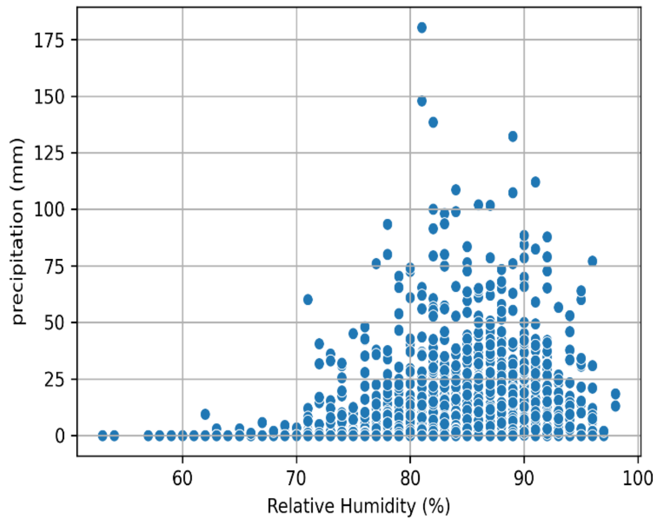

Figure 3. Scatter: Relative Humidity vs Rainfall

Table 3. Statistical Analysis

Condition	Frequency	Average	Median	Minimum	Maximum	Std
No Rain	1129	4.23	4.04	0.51	10.59	1.46
Hujam	1717	2.87	2.74	0.34	8.18	1.08

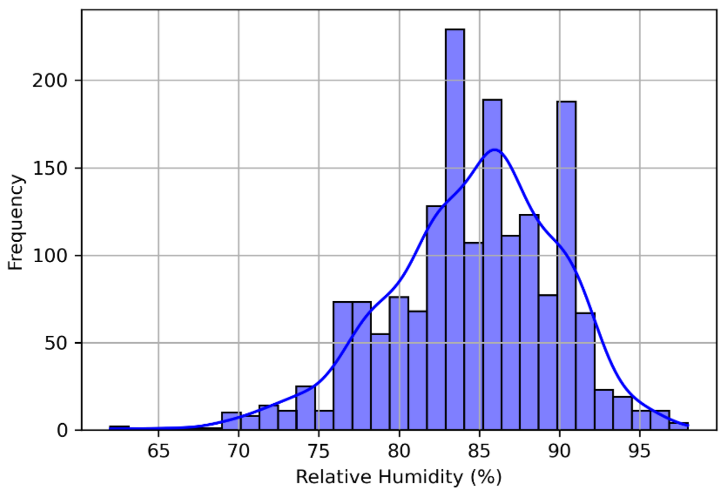

Figure 4. Relative Humidity distribution during rain

Next, to measure how close the air is to saturation conditions, we need the parameter ΔT. Figure 4 shows that ΔT is smaller (average 2.87°C and median 2.87°C) during rain, and the range of ΔT during rain is narrower (0.34°C – 8.18°C). In addition to the relative humidity indicator, ΔT can be an option to analyze the likelihood of rain occurrence. The results of ΔT are more sensitive than humidity in indicating air saturation. Therefore, ΔT can be represented as a margin of unsaturation (Younus Ahmed et al., 1998). In conclusion, a small ΔT is an ideal condition for cloud/rain formation, and if ΔT is large, then the air is dry and there is not enough water for condensation.

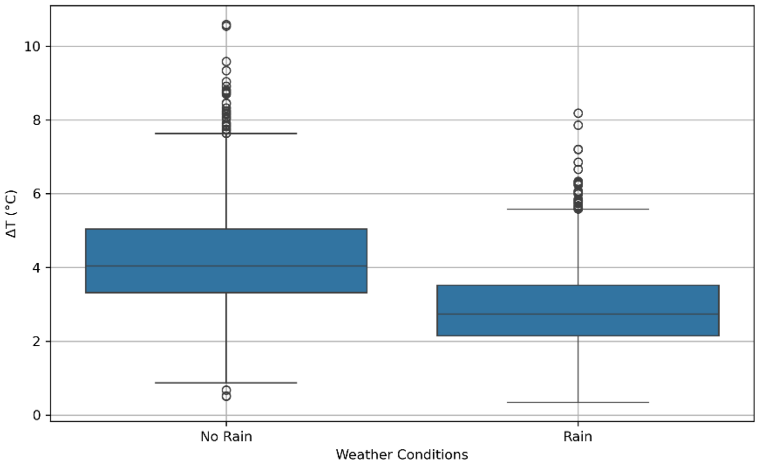

Figure 5. ΔT Distribution during Rain vs No Rain

The collected data is then used to build a machine learning model using the Gradient Boosting Classifier to predict whether it will rain or not. The code is built in several stages, starting with determining the necessary library imports. In this study, the libraries imported are pandas for data manipulation, scikit-learn for modeling and evaluation, and joblib for saving the model. Furthermore, for data preprocessing, rows with missing values are removed, and binary labels are created for rainy and non-rainy conditions.

Table 4. Results from machine learning processing

Gradient Boosting Classifier				
Condition	Precision	Recall	F1-score	Support
0	0.72	0.63	0.67	226
1	0.77	0.84	0.81	343
Accuracy			0.76	569
Macro average	0.75	0.73	0.74	569
Weighted average	0.75	0.76	0.75	569

The data that has been organized includes the variables x, which are temperature, relative humidity, wind speed, month, and day. The variable y is whether it rains or not. The data is then divided into a training set and a test set in an 80:20 ratio. The model that has been built is then evaluated, showing an accuracy of 76% with a recall of 84% for rain. Based on the results in Table 4, this model has several advantages, namely a good ability to predict rain class and consistency against overfitting. The weakness of this model is the imbalance in performance between the two classes. The model appears to overpredict rain conditions due to the imbalance in the number of data points between the rain and no-rain classes. However, this model shows good performance, but it can still undergo refinement.

Ultimately, this research aims to develop a web-based model utilizing machine learning. The web is used to predict rainfall occurrences in the Province of Bandar Lampung. The website is created using the Flask (Python) framework as the back-end and HTML as the front-end interface. The model has been trained and validated in an application that can be accessed and used interactively by users.

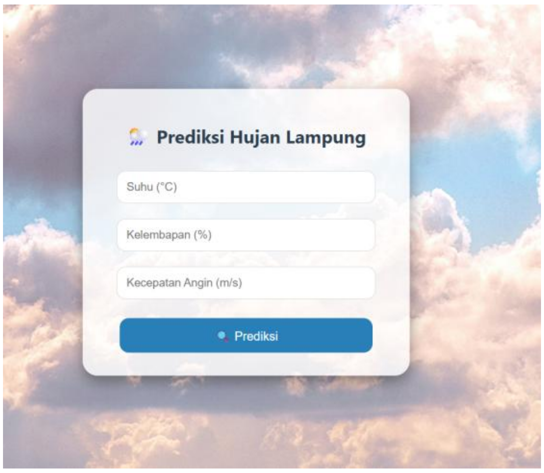

Figure 6. website display

The integration of machine learning models with web applications demonstrates that predictive models are not only theoretical but also practical solutions for society. Websites have practical implications, particularly for the agricultural sector, which is the backbone of Lampung's economy. Farmers can utilize these websites to determine the optimal times for harvesting, fertilizing, and planting crops. The use of the website is expected to minimize losses due to extreme weather.

CONCLUSION

Based on the research, thermodynamic parameters are effective in predicting rainfall. The variables involved include the temperature and dew point difference (ΔT), which is calculated using Magnus' equation as an effective predictor of air saturation conditions that trigger rainfall. Furthermore, a popular machine learning method was used. Integrating these parameters with the Gradient Boosting algorithm resulted in a prediction model with 76% accuracy and 84% recall. These results demonstrate the superiority of a physics-based approach in machine learning. To ensure that this model not only provides computational results but can also be used by the general public, it has been implemented in a ready-to-use web application. Ultimately, it is hoped that this will bridge the gap between atmospheric physics theory and practical needs for early weather warnings, especially for the agricultural sector in Lampung.

Acknowledgments

The author would like to thank the Department of Physics, Faculty of Mathematics and Natural Sciences, University of Lampung, for the facilities and academic support provided during this research. Thanks are also extended to the Energy and Environment Laboratory of the Department of Physics for providing access and resources that supported the smooth running of the research activities.

REFERENCES

Alduchov, A., & Eskridge, R. (1988). *Improved Magnus' Form Approcimation of Saturation Vapor Pressure*. https://doi.org/https://www.osti.gov/servlets/purl/548871&lang=en

Andrews, D. G. . (2010). *An Introduction to Atmospheric Physics*. Cambridge University Press.

Aprilia, A., Wahidin, A. B., & Abdurrahman, A. F. (2025). Integration of Machine Learning and NASA POWER Dataset for Predicting Coffee Production in Lampung. *Jurnal Fisika Flux: Jurnal Ilmiah*

Fisika FMIPA Universitas Lambung Mangkurat, *22*(1), 44. https://doi.org/10.20527/flux.v22i1.20980

Aprilia, A., Wahidin, A. B., & Syafriadi. (2025). Seismic Activity Analysis in Indonesia: Integrating Machine Learning, Geospatial Data, and Environmental Factors for Risk Assessment. *JPF (Jurnal Pendidikan Fisika) Universitas Islam Negeri Alauddin Makassar*, *13*(1), 12–26. https://doi.org/10.24252/jpf.v13i1.54081

Arnah Ritonga, Asni Al Amini, Livia Mutianda, Riamonda Singarimbun, Aiman Hidayat Baeha, Glensius Rayhane Pasaribu, & Juanda Arief Darmawan Damanik. (2025). Analisis Probabilitas Hujan Menggunakan Data Historis Dari BMKG Wilayah I Tahun 2013-2015. *JURNAL RISET RUMPUN MATEMATIKA DAN ILMU PENGETAHUAN ALAM*, *4*(1), 01–20. https://doi.org/10.55606/jurrimipa.v4i1.4367

Catursari, S. A. (2024). *Analisis Pengaruh Tekanan Udara, Kelembapan Udara, dan Suhu Udara terhadap Curah Hujan di Kota Surabaya* [Universitas Islam Negeri Maulana Malik Ibrahim]. https://doi.org/http://etheses.uin-malang.ac.id/65664/1/200604110061.pdf

Chen, P. Y., Chen, C. C., Kang, C., Liu, J. W., & Li, Y. H. (2025). Soil water content prediction across seasons using random forest based on precipitation-related data. *Computers and Electronics in Agriculture*, *230*. https://doi.org/10.1016/j.compag.2024.109802

Halida, I., Sumbri, H., & Dzikrullah, A. A. (2024). Implementasi Uji Mann Whitney Data Pengamatan Automatic Weather Station (AWS) Digi dan Pengamatan Manual di Stasiun Meteorologi Bandar Udara Internasional Juanda Tahun 2021-2022. *Emerging Statistics and Data Science Journal*, *2*(1).

Hikmayanti Handayani, H., Arum Puspita Lestari, S., Cahyana, Y., & Karawang, P. (2025). *Evaluasi Kinerja Algoritma Random Forest Dan Gradient Boosting Untuk Klasifikasi Penyakit Jantung*. *9*(1). https://doi.org/10.31603/komtika.v9i1.13450

Huertas-Abril, C. A., & Palacios-Hidalgo, F. J. (2023). New Possibilities of Artificial Intelligence-Assisted Language Learning (AIALL): Comparing Visions from the East and the West. *Education Sciences*, *13*(12). https://doi.org/10.3390/educsci13121234

Lakshmi, J. V. N. (2018). Machine learning techniques using python for data analysis in performance evaluation. *International Journal of Intelligent Systems Technologies and Applications*, *17*(1/2), 3. https://doi.org/10.1504/ijista.2018.10012853

Markovics, D., & Mayer, M. J. (2022). Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. *Renewable and Sustainable Energy Reviews*, *161*. https://doi.org/10.1016/j.rser.2022.112364

Ningsih, N., Iskandar, N. A. S., Rizqiyah, S., & Sudriyanto, S. (2024). PREDIKSI CHURN PELANGGAN INDUSTRI TELEKOMUNIKASI MENGGUNAKAN METODE ARTIFICIAL NEURAL NETWORK BERBASIS STREAMLIT. *JUSTIFY : Jurnal Sistem Informasi Ibrahimy*, *3*(2), 105–114. https://doi.org/10.35316/justify.v3i2.5544

Prakasa, A., & Utami, F. D. (2019). Sistem Informasi Radar Cuaca Terintegrasi BMKG. *Journal of Telecommunication, Electronics, and Control Engineering (JTECE)*, *1*(02), 78–87. https://doi.org/10.20895/jtece.v1i02.89

Schneider, T., O'Gorman, P. A., & Levine, X. J. (2010). Water vapor and the dynamics of climate changes. In *Reviews of Geophysics* (Vol. 48, Issue 1). https://doi.org/10.1029/2009RG000302

Wang, H., Zhang, J., Chen, L., & Li, D. (2022). Relationship between summer extreme precipitation anomaly in Central Asia and surface sensible heat variation on the Central-Eastern Tibetan Plateau. *Climate Dynamics*, *59*(3–4), 685–700. https://doi.org/10.1007/s00382-022-06148-w

Wang, W., & Hocke, K. (2022). Atmospheric Effects and Precursors of Rainfall over the Swiss Plateau. *Remote Sensing*, *14*(12). https://doi.org/10.3390/rs14122938

Wang, W., Huang, D., Wang, X. G., Liu, Y. R., & Zhou, F. (2011). Estimation of soil moisture using trapezoidal relationship between remotely sensed land surface temperature and vegetation index. *Hydrology and Earth System Sciences*, *15*(5), 1699–1712. https://doi.org/10.5194/hess-15-1699-2011

Younus Ahmed, S., Gandhidasan, P., & Al-Farayedhi, A. A. (1998). PIPELINE DRYING USING DEHUMIDIFIED AIR WITH LOW DEW POINT TEMPERATURE. In *Applied Thermal Engineering* (Vol. 18, Issue 5).